

Application of data mining technology in medical image processing

Rui Wang ^{1,a}, Jinguo Wang ^{2,b*}, Na Wang ^{3,c}

¹Department of Information Engineering, Jilin Business and Technology College
China

²Department of Urology, the First Hospital of Jilin University, China

³Department of Anaesthesiology, the First Hospital of Jilin University, China

^aXiaoben6666@126.com, ^bwangjinguolily@163.com, ^clilyly12345@163.com

*corresponding author

Keywords: Data mining. Medical image processing.

Abstract. Medical image research and application has drawn attention from many research fields, and the joint efforts of these fields have made the research and extension of the research, and made great contributions to the cause of human health. This thesis focuses on the theory and practice of data mining, and analyzes the data mining problem of medical image, and comprehensive existing medical image feature system, based on the idea of pixel density, grid and image classification.

1. Introduction

Data mining technology has been developed rapidly since the last century 90 's, and has gradually formed a unique data integration, knowledge mining, pattern evaluation and verification theory and method system. Data mining plays an important role in many fields such as finance, manufacturing, telecommunications, insurance, etc.. The multimedia data such as image, audio, video and text are accumulated, and the data mining technology is put forward higher request[1].

Image mining is an important part of multimedia data mining. The research content is the subject of computer vision, image processing, data mining, machine learning, database and artificial intelligence[2]. Specifically, image mining is a feature vector which can represent the structure of the image from the image, and the distance between them is analyzed. Through the analysis of the content of the image, indexing, classification and retrieval, and other operations, to further find the knowledge or patterns of interest[3].

Image mining is not a simple application of data mining technology in the field of image, because the existing data mining technology is mainly applied to structural data, while the image data is nonstructural data. Data in structural data have a certain meaning, while the data in the image data does not necessarily have a clear semantics. In order to meet the needs of mining knowledge in the image database, some new mining algorithms and other related technologies are needed[4].

Medical image has its distinctive features. One is the gray level resolution. Medical image gray-scale resolution several times in ordinary gray image, for medical image, color feature of ordinary gray image is no longer applicable; the second is computer image reconstruction. The imaging principle of CT, MRI and other medical images in the medical field is based on the density difference of human body. After the reconstruction of the human body image data matrix sequence, it is different from the ordinary gray image; three is the objective expression of human anatomy. Medical image has its specific meaning, and the image information is unified with medical semantics[5].

Research on Data Mining Based on medical image has just started, and there are many problems need to be studied and discussed. First of all, the characteristic data for medical image mining is limited to the color, texture, shape and so on. Second, the medical image data is a complex and high dimensional data mining algorithm, which is used in medical image mining. The medical image data is a complex and high dimensional data mining algorithm. Third, the data mining of medical images is limited to individual specific tissues and organs, such as breast, brain and so on, and the research results are not generally applicable to other organs. Fourth, the limitation of the existing research on the global features of the whole image of mining research, and clinical medicine to human body image local features is more clinical significance. These problems and limitations are the starting point and the main content of this thesis. It is obvious that the research of medical image data mining technology has an important role in promoting the development of digital medical technology. It provides technical support for the objective, scientific and accuracy of medical image diagnosis[6].

2. Image data mining

Image data mining is an important part of multimedia data mining. It is the hidden knowledge in the image data collection, including the relationship between the image, the relationship between the image and the character data, the relationship among the entities in the image, and other patterns or relations. It is a cross field of computer vision, image processing, image retrieval, data mining, machine learning, database and artificial intelligence. In particular, image data mining is the feature vector extracted from the image to distinguish the structural content of the image, in which the space of the feature vector is compared and analyzed. The feature vectors of image can be divided into two types, the feature vector of the whole image content feature vector and characterization of the local image content expression. Image mining is not a simple application of data mining technology in the field of image, because the existing data mining technology is mainly applied to structural data, while the image data is nonstructural, structural data have a certain semantic, while the image data is not necessarily a significant semantic. In order to meet the needs of mining knowledge from the image database, it is necessary to develop some new algorithms and other related technologies.

3. Medical image data mining

To study the characteristics of medical images and to explore the suitable mining algorithm for medical image data is the two key issues of image mining. Image information is very difficult to use the only description of human language, usually based on the characteristics of content, such as color, texture, shape, spatial relations and other low-level physical features, including a certain knowledge background, such as round, square, clustering, etc., and these features can not only represent the real content of the image, and can use the machine automatically to overcome the shortcomings of manual extraction of image features and time consuming. The algorithms used in image mining include decision tree, clustering analysis, association rules, neural network, etc..

Medical images have the characteristics of high gray scale, obvious meaning, easy reconstruction and so on. The characteristics of medical image are closely related to the characteristics of medical image, and the characteristics of medical image mining are limited to some basic color, texture,

shape and so on. The medical image data mining algorithm is still in the direct use of traditional data mining algorithms such as decision tree, neural network and association rules. The medical image data is complex and multi-dimensional, and it needs to be studied.

We define the features of medical image, gray scale and its density distribution, study the data mining algorithm based on image feature extraction, explore the gray level and its density distribution and clustering analysis of normal human tissues and organs.

Object recognition. Object recognition is a hot research in the field of image processing. This technique is used to recognize the objects in the real world. Recognition image is one of the main tasks of image mining. Automatic machine learning and extracting meaningful information can only be realized when a number of objects are recognized and identified by the machine. Object recognition is the process of identifying the object of the real world.

A typical object recognition system consists of four parts, i.e., the model database, feature vector, assumptions and hypothesis verifier. The model database contains all the known models in the system, which includes the important features of the object. The original feature of the image is detected by using these at the pixel level, which can help the hypothesis to be given to the possible eigenvalues of the object in the image. Assuming the verifier to verify the hypothesis by using the known model and exact expression may feature object value. The system ultimately selects the most likely targets in the relevant object.

Image retrieval. Image mining needs to carry out the image according to a certain requirement to carry out the standard search. Image retrieval can be based on different levels, such as pattern and knowledge level, semantic concept layer, object layer, pixel layer, and so on.

Three kinds of query patterns are proposed for image retrieval: retrieval based on association rules, based on the description of image retrieval and content-based image retrieval. Retrieval based on correlation property, it is necessary to adjust the traditional table structure to meet the need of image. This retrieval model still needs to be organized, and the image is extended into a special field, but when there is a large bitmap, it includes image or graphic file. This search is similar to the search operation based on the text, but it is extended to the image data. The result of the retrieval may be the property of the image to meet the requirements of the relevant information, of course, sometimes may also meet other related properties. Based on the description of the retrieval, the basic idea of the retrieval is to locate the interested images through the image storage description and user image retrieval requirements. The image description is often referred to as the tag or keyword, which is done by manual operation, and is assigned to each image in the image preprocessing stage to mark or key words. With the appearance of large scale image database, the problem of "vocabulary" and "no certainty" is more and more obvious. Therefore, the content-based image retrieval is proposed to overcome these problems. Content based image retrieval has achieved many results in the extraction of visual information, image indexing and retrieval system.

Image index. Current image retrieval system is based on the similarity measurement method, and the method based on Euclidean distance is sometimes applied. Similarity measurement is realized by using the histogram intersection, cosine, correlation, etc.. An effective method is to reduce the dimension and then use appropriate multidimensional indexing techniques, which can support non Euclidean similarity measure method. The image retrieval system has been developed based on the Oracle platform, which can be used to operate on the high dimensional approximate data. These data can represent the image, reduce the search space and reduce the computational cost. Haykin S. describes a new image indexing technique that is compressed by using the compressed image

features as the key to retrieve the image, and the various indexing methods are focused on the specific image features. Lin K. gives an effective color indexing method, which is based on the similarity search, and the search time increases with the size of the database. L. Tan K. proposed a multi-level R tree index, called nested R tree search, greatly improved the efficiency of the search. With the increase of the mechanism of image indexing, L. Tan K. gives a color space index technology, which is a method to evaluate the performance of spatial index technology.

References

- [1] U Krohn,N J Davies,R Weeks. Concept Lattices for Knowledge Management[J]. BT Technology Journal . 1999 (4).
- [2] Derrick G. Kourie,G. Deon Oosthuizen. Lattices in machine learning: Complexity issues[J]. Acta Informatica . 1998 (4).
- [3] S. Colantonio,O. Salvetti,I. B. Gurevich,Yu. Trusova. An ontological framework for media analysis and mining[J]. Pattern Recognition and Image Analysis . 2009 (2).
- [4] Wynne Hsu,Mong Li Lee, Ji Zhang. Image Mining: Trends and Developments[J]. Journal of Intelligent Information Systems . 2002 (1).
- [5] Liu Guili,Li Chanliang. Vector Chinese characters database and application to the embedded marking machine systems[A]. Proceedings of the Fourth International Symposium on Test Automation & Instrumentation[C].